DOCUMENT RESUME

ED 445 031 TM 031 636

AUTHOR Berkovits, Ilona; Hancock, Gregory R.
TITLE A Comparison of Methods for Structural Modeling with Polytomous and Continuous Variables.
SPONS AGENCY Educational Testing Service, Princeton, NJ.
PUB DATE 2000-04-00
NOTE 33p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Computer Software; *Estimation (Mathematics); Simulation; *Structural Equation Models
IDENTIFIERS Continuous Variables; *LISREL Computer Program; *Parameter Identification; Polytomous Variables

ABSTRACT

A simulation study compared three methods of estimating parameters within structural equation models (SEM) with polytomous variables. These methods appear in three SEM computer software packages: (1) LISREL (Joreskog and Sorbom, 1996) with PRELIS (Joreskog and Sorbom); (2) EQS (Bentler, 1995); and (3) the new Mplus (Muthen and Muthen, 1998). The differences in parameter estimation accuracy were evaluated by manipulating sample size, exogenous and endogenous factors with polytomous indicator variables, number of polytomous variables per factor, number of categories per polytomous variable, and different thresholds for the underlying continuum to create the categories. Results show some severe parameter estimation problems in LISREL (LISREL results are not discussed in this paper). EQS had problems with convergence, but the parameter estimates did not appear to be affected, allowing comparison of EQS results with Mplus results. Mplus displayed a very slight estimation difference depending on whether the scaling variable on the endogenous factor was polytomous. Mplus did appear to estimate the path coefficients closer to the true values than EQS for all conditions, with Mplus results having less variability. Mplus is faster and converges more often than EQS, and thus appears preferable for the estimation of SEM models with polytomous variables. (Contains 8 tables, 4 figures, and 19 references.) (Author/SLD)

A comparison of methods for structural modeling with polytomous and continuous variables

Ilona Berkovits and Gregory R. Hancock

University of Maryland, College Park

Correspondence concerning this paper should be directed to

Ilona Berkovits
1230 Benjamin Building
University of Maryland
College Park, MD 20742-1115
phone: (301)405-3623
fax: (301)314-9245
e-mail: ilona@wam.umd.edu

This research was supported by the Gulliksen Psychometric Fellowship from the Educational Testing Service (ETS).

Abstract

The purpose of this simulation study is to compare three methods to estimate parameters within structural equation models with polytomous variables. These methods appear respectively in three SEM computer software packages: LISREL (Jöreskog & Sörbom, 1996b) together with PRELIS (Jöreskog & Sörbom, 1996a); EQS (Bentler, 1995); and the new Mplus (Muthén & Muthén, 1998). The differences in parameter estimation accuracy among the estimation procedures are evaluated by manipulating several conditions which include: sample size, exogenous and/or endogenous factors with polytomous indicator variables, number of polytomous variables per factor, number of categories per polytomous variable, and different thresholds (cutoff points) on the underlying continuum to create the categories on the polytomous variable(s).

Unfortunately, the results show some severe parameter estimation problems in LISREL, which requires further investigation, and the results from the LISREL program are not discussed here. EQS also had problems, although these were with convergence. The parameter estimates did not appear to be affected by the convergence problems, and therefore EQS results are compared to the Mplus results. Mplus displayed a very slight estimation difference depending on whether or not the scaling variable on the endogenous factor was polytomous. Still, Mplus appears to estimate the path coefficients closer to the true values than EQS for all conditions, with Mplus estimates also having less variability. In addition, Mplus is faster, and converges more often than EQS. For these reasons, Mplus appears to be preferable to EQS for researchers performing estimation of SEM models with polytomous variables.

A comparison of methods for structural modeling with polytomous and continuous variables

Traditionally, structural models have assumed that the observed variables are continuous with a multivariate normal distribution. However, in the social sciences, variables that have an underlying continuum are often measured using dichotomous or polytomous scales. Therefore, models with these observed categorical variables may not be estimated properly using traditional methods. Several methods have emerged in structural equation modeling (SEM) computer software that attempt to deal with combinations of polytomous and continuous variables. These methods are still predicated on (conditional) multivariate normality, however this assumption applies to the distribution of the assumed underlying continuum for the polytomous variables, rather than the categorical variables themselves.

A simulation study is used to evaluate the adequacy of three methods to estimate models with polytomous variables. These methods appear respectively in three SEM computer software packages: LISREL (Jöreskog & Sörbom, 1996b) together with PRELIS (Jöreskog & Sörbom, 1996a); EQS (Bentler, 1995); and the new Mplus (Muthén & Muthén, 1998). The estimation process differs in each software package, and these estimation methods are briefly described later in this paper.

The differences in parameter estimation accuracy among the estimation procedures are evaluated by manipulating several conditions which include: sample size, exogenous and/or endogenous factors with polytomous indicator variables, number of polytomous variables per factor, number of categories per polytomous variable, and different thresholds (cutoff points) on the underlying continuum to create the categories on the polytomous variable(s).

Background

Assumptions

Assumptions in SEM include that the observed variables are conditionally multivariate normally distributed. An additional assumption, that all relations among variables are linear, is probably reasonable when the variables satisfy distributional assumptions. Muthén (1993) clarifies that these assumptions should hold for the latent variables underlying categorical variables.

While regression has numerous diagnostics to evaluate distributional assumptions for a regression model, Bentler (1995) states that, at present, there are no good corresponding diagnostics in SEM. Muthén (1993) suggests several tests to assess bivariate normality, but concludes that a very large sample size is needed to perform them, and that "rejections of the normality model are frequently found" (p. 217). No recommendations are given on how to proceed if the normality model is rejected.

Another traditional assumption in SEM is that variables are continuous. In sample data, continuity will never be observed because the largest number of different scores that could be obtained is the number of subjects in the study. However, many variables, income for example, can be seen as continuous in theory. In practice, continuous variables may be categorized for convenience; for example, income may be recorded as one category if it is less than $5,000, as another category if it is between $5,000 and $15,000, and so on. These are categorical scores that have an underlying continuum. Therefore, an assumption when performing an analysis with polytomous variables in SEM is that a continuum underlies the variable, as well as that the continuum itself satisfies the aforementioned distributional assumptions. Note that variables that are intrinsically categorical, like gender, do not contain an underlying continuum, and do not

satisfy these assumptions. Methods to deal with intrinsically categorical variables are not included in this paper.

Correlations with Polytomous Variables

Because Pearson product moment correlations assume linearity and continuity, simply using Pearson correlations would lead to distorted estimates of polytomous variables' relationships with other variables (Muthén & Kaplan, 1985; Olsson, 1979). One approach to remedy this problem is to map a categorical variable nonlinearly into a linear continuous variable, and then develop linear structures (correlations) for the new continuous variable. This is the basis for tetrachoric, polychoric and polyserial correlations (e.g., Lee & Poon, 1986).

Tetrachoric correlations estimate the relation underlying pairs of binary variables, while polychoric correlations estimate the relation underlying pairs of polytomous variables with any number of categories on each. Polyserial correlations estimate the underlying relationship of polytomous variables paired with continuous variables. In these estimations, thresholds need to be estimated for the categorical variables. The thresholds are the cutoff values on the underlying continuous variables presumed to create the observed categorical variable. These thresholds are often given in z-score form.

Many maximum likelihood methods exist to estimate the polychoric and polyserial correlations, and the thresholds (eg. Lee, Poon & Bentler, 1992; Muthén, 1984). Some methods propose to estimate all the correlations simultaneously – including the correlations among continuous variables. Others break the estimation into stages. Many different multi-stage estimation procedures have been proposed. For example, one simple procedure could be to estimate the continuous variables' correlations in one stage, estimate the polyserial correlations in a second stage, and estimate the polychoric correlations in a third stage - with the correlations

estimated simultaneously within each stage. Another even simpler method could be to break this process down into its smallest components, and estimate each pairwise correlation, regardless of type, by using only the information from the pair. On the opposite end of the spectrum, there have been methods proposed which estimate all parameters in a model (the correlations, thresholds and the model paths and loadings) simultaneously (Lee, Poon & Bentler, 1992). This has been referred to as the one-stage full maximum likelihood model (FML), but is currently considered too complex for practical parameter estimation and does not appear in accessible software packages. Three different methods, which are less complex, but attempt to use as much information as possible in estimation, and are available in popular SEM software packages, are described in the following section. The procedures for estimating the underlying correlations are emphasized, as the estimation of structural models is based on correlations, therefore one would assume that the method which best estimates these correlations would also provide the best estimate of the structural model parameters underlying the data.

Methods for Estimating Models with Polytomous Variables in Three SEM Software Packages

EQS (Bentler, 1995; Bentler & Wu, 1995)

Estimation in EQS, when including polytomous variables in the structural equation model, occurs in two major stages: first all the correlations and thresholds are estimated, then the basic parameters of the model are estimated from these correlations using an arbitrary generalized least squares (AGLS) procedure. The first stage is further broken down into two parts: in the first part, the correlations of the continuous variables with each other are estimated simultaneously with the polyserial correlations; in the second part, only the polychoric

correlations are all estimated simultaneously. This is referred to as the partition maximum likelihood procedure (PML) (Lee, Poon & Bentler, 1995).

In estimating the polyserial correlations separately from the polychoric correlations, the thresholds are estimated twice for each polytomous variable (once for each part), and therefore different threshold estimates may be found. Lee, Poon and Bentler (1995) do not consider this a serious issue, stating that the difference in these threshold estimates are often tiny and that the thresholds themselves are not used in estimating the rest of the model. After the correlations are estimated in the first stage, an AGLS estimation method is used to estimate the model parameters in the second stage.

Requirements to run a model with polytomous variables in EQS include: raw data must be analyzed; polytomous variables should not have more than 5 to 7 categories; and a reasonably large sample size is recommended. The reason for these requirements is that the categorical data methods require the cross-tabulation of the categorical variables. The more categories and the smaller the sample size, the more likely you are to have empty cells in these cross-tabulations, which leads to a breakdown in the computational procedures (Bentler, 1995).

PRELIS (Jöreskog & Sörbom, 1996a) and LISREL (Jöreskog & Sörbom, 1996b);

The procedure used by these two programs together first uses PRELIS to obtain the polychoric and polyserial correlation estimates, and an asymptotic covariance matrix. Then LISREL estimates the basic parameters of the model from these correlations using a weighted least squares procedure, where the weight matrix is the inverse of the estimated asymptotic covariance matrix **W** of the polychoric and polyserial correlations (Jöreskog & Sörbom, 1988b). This differs from EQS in that the estimation of the polychorics and polyserials are "estimated from the observed pairwise contingency tables of the ordinal variables" (p. 193). While EQS

uses some simultaneous estimation procedures to estimate these correlations, PRELIS/LISREL estimates the correlations using only information from two variables at a time. Thresholds are estimated univariately from the (marginal) distribution of each variable.

Mplus (Muthén & Muthén, 1998)

Mplus is based on Muthén's (1984) proposed three-stage estimation method, with the third stage modified by Muthen, du Toit, & Spisic (in press). In the first stage, this method estimates the relationships between variables in terms of correlations and thresholds, using procedures similar to those used in PRELIS/LISREL. The second stage consists of estimating the weight matrix based on these results. Finally, the third stage estimates the model parameters with a robust weighted least-squares (WLS), by using the diagonal of the estimated weight matrix, and using the full weight matrix for robust standard errors and Chi-square. The parameters are simultaneously estimated within this third stage.

Simulation Study

The simulation is designed to investigate the adequacy of parameter estimates and fit statistics for a known population model as obtained by three estimation procedures contained in the following structural equation modeling computer software packages: LISREL (Jöreskog & Sörbom, 1996b) together with PRELIS (Jöreskog & Sörbom, 1996a); EQS (Bentler & Wu, 1995); and the new Mplus (Muthén & Muthén, 1998). The effect of polytomous variables (with a known underlying continuum) on structural model parameter estimation is of primary interest in this study.

These effects are investigated using a single population model with two exogenous factors, one endogenous factor, and three indicator variables per factor. As shown in Figure 1,

the true values for the factor loadings of variables V1 through V9 are all equal to .8, the path coefficient F3,F1 equals .5, and the path coefficient F3,F2 equals .7. This leads to a disturbance of .51 on F3, and error terms of .6. These population values were set as start values for parameter estimation in the three software packages. For identification purposes, the variances of the exogenous variables were set to 1, and the factor loading V7,F3 was set to the population value of .8.

Two-hundred datasets, with nine continuous variables, were generated using EQS, based on the population model values shown in Figure 1, for each of the following sample sizes: 100, 200, 500, 1000, 5000 (for a total of 1000 original datasets with continuous variables). For each condition, designated variables are categorized at specified threshold values, as outlined in Table 1, within the following factorially-crossed conditions: factors containing polytomous indicators (1 exogenous factor; 2 exogenous factors; 1 endogenous factor; one combination of 1 endogenous and 1 exogenous; and all factors), number of polytomous indicator variables on a factor (1, 2, 3), and number of categories per polytomous variable (2, 3, 5, 7).

The threshold values for polytomous variables, outlined in Table 1, are manipulated for each level of number of categories. The different threshold values result in different proportions of observations in each category — equal proportions in each category (uniform), unequal proportions in each category (large proportion in middle categories (normal), large proportion in outer categories (U-shaped), large proportion in low categories (positively skewed). Therefore this condition is nested within number of categories per polytomous variable. The threshold values used to create the polytomous variables from continuous variables are given in z-score units, and the corresponding percentage of cases in each category for the categorized normally distributed underlying continuous variable is additionally presented in Table 1. This design

results in a total of 1,050 cells, and the categorical variables' underlying continuums are normally distributed for all conditions.

This study compares the relative accuracy of the three outlined estimation procedures for model parameters under various conditions. In addition, the relationship of the polytomous variable conditions to model parameter estimation is of interest, and will be evaluated based on comparisons of the estimated model parameters to the known population parameters. Findings should lead to methodological and practical recommendations for SEM researchers using polytomous variables.

Results

Estimation Problems

Problems with convergence in EQS

As previously outlined, categorical data methods in EQS require the cross-tabulation of the categorical variables. The more categories and the smaller the sample size, the more likely you are to have empty cells in these cross-tabulations, which leads to a breakdown in the computational procedures (Bentler, 1995). This result is seen in Table 2, where out of 200 replications, the overall average number that converged can be seen to be decreasing from just over 146 with 2 categories, to only just above 84 with 7 categories per variable, even with the population values used as start values. There appears to be a relationship with sample size since the average number of converging runs increases as sample size increases, and as the number of categories per variable decreases. However, a surprising result is that the standard deviation of the mean number of converging replications increases as the sample size increases, except when number of categories equals 2. This may be due to some cells with large sample sizes requiring

increased memory or computations with a large sample size that may have strained either the EQS program's capacity or the computer's capacity, and therefore failed to converge in some of those instances.

Although EQS appears to struggle with converging, Mplus converged for all 200 replications under all conditions, and LISREL converged over 99% of the time.

While EQS' struggle to converge appears to be related to number of categories per variable, the parameter estimates for the converging runs do not seem to be related to the number of categories per variable, nor does there appear to be any obvious relationships between the size of the parameter estimates to any of the other design factors. This will be presented in the next section.

Problems with parameter estimation under certain conditions with LISREL

The average LISREL estimate for the factor loadings of variables V4 through V9 all fell between .7 and 1.0, as compared to the population value of .8. However the average estimate for the factor loadings of V1, V2, and V3 had many high values exceeding 1.0, and when these occurred, the estimates for the path coefficients between factors were also seriously affected (see Table 3). There appears to be a strong relationship between which factors have all variables categorized, and the problems with estimation. Figure 2a clearly shows the pattern of overestimation of the F3,F1 path when F1 and F3's variables are all categorized. While this may appear to make some sense at first, this does not explain why the F2,F3 path is severely underestimated when the same F1 and F3 variables are categorized as seen in Figure 2b. Unlike the other conditions, under this condition parameter estimates do not appear to converge to the true value at n=5000. It is also not clear why when we have all three factors with all variables categorized, the estimation is nearly perfect, averaging .501 across all sample sizes, when having two out of the three presented such a problem.

Due to the severity of this problem with parameter estimation in LISREL under the condition mentioned, the results from the LISREL portion of the simulation will not be discussed further until the source of the error can be determined as programming error (either in the simulation or in the software package) or if it is due to estimation method. To discuss the average performance of LISREL estimates, with the extreme values possibly canceling each other out, as compared to the other programs that do not appear to have such extreme mean values, seems inappropriate. Although LISREL may have satisfactory and comparable results to EQS and Mplus when the variables on F1 are not all categorized, to discuss these 'good' results without determining the source of the problems in the other conditions may be misleading.

Differences in parameter estimation under certain conditions with Mplus

For smaller sample sizes Mplus has a larger range of estimates of variable's factor loadings when F3 contains categorized variables, although on average, these estimates are very close to the population value of .8. The increased range of values may be due to the scale for F3 being determined by the loading for V7, which is categorized in this condition. However, when F3 does not contain any categorized variables, the range of values decreases, but the estimates are slightly overestimated on average (see Figure 3a for V9 example). Estimates of factor loadings on other variables are slightly closer to the true value for variables in conditions where the factor they load on contains at least one categorized variable. Both path coefficients are slightly overestimated in conditions where V7 is a categorized variable, which is the scale factor for F3, the exogenous factor (Figure 3b shows the effect for the F3,F1 path).

Differences in Path Coefficient Estimation in Mplus and EQS (converging runs only)

Under Various Conditions

Sample size (100, 200, 500, 1000, 5000)

Five sample sizes were tested in this simulation: 100, 200, 500, 1000, and 5000 subjects. Mplus converged for all 200 replications for all sample sizes and all conditions. For sample sizes of 100, EQS converges about 48 times out of the 200 replications, and for sample sizes of 5000, increases to converging an average of about 136 times, out of 200 replications, as shown in Table 2.

Mplus appears to estimate both path coefficients closer, on average, to the true values than EQS estimates them, for all sample sizes (see Figure 4). Table 4 shows that for the F3,F1 path, which has a population value of .5, EQS's mean estimate with n=100 is .476 (.024 below the true value), while Mplus's estimate is .511 (.011 above the true value). Both converge, on average, to almost exactly .5 with n=5000, with the mean EQS estimate of .5008 and the mean Mplus estimate of .5002. Note however that while EQS's mean estimates at n=5000 had a range of .018, Mplus had a much smaller range of only .003. In addition, Mplus's standard deviations of the original estimates across all replications are smaller, on average, than EQS's. A similar pattern of results, closer estimates by Mplus with a smaller range (see Table 4) and smaller average standard deviation (see Figure 5) than EQS, is found for the F3,F2 path coefficients as well.

Number of categorized variables (1, 2, 3, 4, 6, 9)

Due to the design of the simulation, six different numbers of categorized variables occured: 1, 2, 3, 4, 6, and 9. For most of these cases, Mplus appears to overestimate the path coefficients, while EQS tends to underestimate them on average. However, Mplus values are

closer, on average, to the true values than the EQS estimates (see Table 5). For the F3,F1 path, which has a population value of .5, EQS' mean estimate when only one variable is categorized, is .478 (.022 below the true value), while Mplus' estimate was .502 (.002 above the true value). However, EQS and Mplus' estimates are about equidistant from the population value of .5 for the F3,F1 path when four or six variables are categorized, although Mplus tends to overestimate slightly (with a mean estimate of .502 and .503 respectively) , while EQS tends to underestimate slightly (with a mean estimate of .496 and .497 respectively).

While Mplus' mean estimates are consistent across all numbers of categorized variables, EQS' estimates seemed to get closer to the true value as the number of categorized variables is increased, except for when all the variables were categorized. While on average Mplus consistently overestimates by .001 to .003 over the true value, EQS' mean estimates are less consistent under these conditions, and underestimated from a small .003 to a larger .022.

In addition Mplus had much smaller ranges, close to around .03 for every number of variables categorized, while EQS' overall mean estimates had a range of up to .11 when all nine variables are categorized. The same pattern of results, closer estimates by Mplus for each specific number of categorized variables, and with a smaller range than EQS, is found for the F3,F2 path coefficients as well (see Table 5).

Pattern of factors containing polytomous indicators

The design of the investigation contained 5 different configurations of factors which contain polytomous indicators: 1 exogenous factor; 2 exogenous factors; 1 endogenous factor; one combination of 1 endogenous and 1 exogenous; and all factors. Mplus appears to best estimate the path coefficients when F3, the exogenous factor, does not contain a categorized variable, and estimates them nearly perfectly to 3 decimal places on average (see Table 6)

In all five conditions, Mplus estimates the F3,F1 path coefficients closer, on average, to the true values than EQS. Mplus tends to be very close to the parameter values on average, while EQS tends to slightly underestimate them. Mplus had a much smaller range overall of about .031, while EQS mean estimates ranged by up to .149 overall when all factors contained categorical values. A similar pattern of results is found for the F3,F2 path coefficients as well (see Table 6).

Categories per polytomous variable (2, 3, 5, 7)

Thresholds outlined in Table 1 categorized the continuous variables into polytomous variables with either 2, 3, 5, or 7 categories. Mplus appears to overestimate the path coefficients closer, on average, to the true values than EQS is under-estimating them, for all numbers of categories per polytomous variable (see Table 7). For the F3,F1 path which has a population value of .5, EQS' mean estimate with only 2 categories per variable, is .486 (.014 below the true value), while Mplus' estimate was .502 (.002 above the true value). In the conditions when there are 3 categories per variable, EQS' best estimates on average occur, and Mplus' worst estimates occur. Even so, the mean EQS estimate of .488, is still .012 below the true value, while the mean Mplus estimate of .503 is only .003 above the true value. Again, Mplus had much smaller ranges of up to .03 with two and three categories per variable, while EQS mean estimates ranged by up to .149 with 2 categories per variable. A stronger, but similar result occurs for the F3,F2 path. However, there is no clear trend to indicate a linear relationship between number of categories per variable and estimation by either program's method.

Type of threshold categorization (uniform, normal, U-shaped, positive skew)

Thresholds outlined in Table 1 categorized the normal continuous variables into polytomous variables with different proportions of cases in each category, to create different

distributional appearances of the categorized variables. The thresholds were manipulated to have the categorized variables approximate uniform, normal, U-shaped, and positively skewed distributions. Mplus appears to estimate the path coefficients closer, on average, to the true values than EQS estimates them, for all categorized distributions (see Table 8). For the F3,F1 path which has a population value of .5, EQS's mean estimate across all categorical distribution types is .487 (.013 below the true value), while Mplus's mean estimate is .502 (.002 above the true value). Again, the range of mean estimates is larger for EQS (up to .149 in the positively skewed condition) than for Mplus (up to .031 in the positively skewed condition). Both seem to have fairly consistent estimates across all different types of categorizations. As both programs assume that the underlying variable is normally distributed, and this assumption holds in this simulation, it appears that each program is consistent in estimating the parameter values regardless of the threshold's locations to categorize the underlying continuous variable. The same pattern of results, closer estimates by Mplus with a smaller range than EQS, is found for the F3,F2 path coefficients as well (see Table 8).

Discussion

The purpose of this simulation study is to compare three methods to estimate structural equation models with polytomous variables. These methods appear respectively in three SEM computer software packages: LISREL (Jöreskog & Sörbom, 1996b) together with PRELIS (Jöreskog & Sörbom, 1996a); EQS (Bentler, 1995); and the new Mplus (Muthén & Muthén, 1998).

Unfortunately, the results for LISREL were not included due to the severe parameter over- and under-estimation problems in LISREL when the indicator variables on F1 and F3 are all categorized. This requires further investigation to determine if the source of the error is due

to programming error (either in the simulation or in the software package) or if it is due to the estimation method. In these cases, values estimated were so extreme, that the results from the LISREL program will not be discussed, even though other cases had acceptable estimates. To do so might mislead researchers about the performance of estimation with the LISREL program when it is not clear what is causing major estimation difficulties.

EQS also had problems, although these were with convergence. The parameter estimates did not appear to be affected by the convergence problems, and therefore EQS results are compared to the Mplus results.

Mplus did not appear to have any problems with convergence. The only evidence of difficulties was a very slight overestimation when the endogenous factor, F3, contains categorized variables. This may be due to the scale for F3 being determined by the loading for V7, which is categorized in this condition.

Methodologically, EQS uses more simultaneous estimation of variable's relationships (correlations and thresholds) than Mplus. However, Mplus' parameter estimation using a robust WLS procedure appears to compute superior parameter point estimates, as well as having smaller variability, than EQS' estimates which were computed using AGLS methods.

In conclusion, Mplus appears to generally estimate the path coefficients closer to the true values than EQS for all conditions included in this study (on average, Mplus has a tendency to slightly overestimate the parameters, while EQS has a tendency to underestimate them). In addition, Mplus' ranges of the mean values, and standard deviations of the raw parameter estimates are smaller, on average, than EQS'. Mplus estimation is also faster, and converges more often than EQS. For these reasons, Mplus appears to be preferable to EQS for researchers performing estimation of SEM models with polytomous variables.

Further research is recommended to compare the performance of these different estimation methods when the continuums underlying the polytomous variables are not normally distributed. In addition, the effect of crossing different numbers of categories per polytomous variable across one factor, and across several factors, may be looked into. More complex structural models can, of course, also be devised.

References

Bandalos, D.L., & Enders, C.K. (1996). The effects of heterogeneous item distributions in confirmatory factor analysis using ML and WLS estimators. Paper presented at the annual meeting of the American Educational Research Association, New York, N.Y.

Bentler, P.M. (1995). EQS structural equations program manual. Encino, CA: Multivariate Software, Inc.

Bentler, P. M. & Chou, C. P. (1988). Practical issues in structural modeling. In J. S. Long (Ed.), Common problems/Proper solutions. Beverly Hills, CA: Sage Publications.

Bentler, P. M. & Wu, E. J. C. (1995). EQS for Windows: Users guide. Encino, CA: Multivariate Software, Inc.

Jöreskog, K.G., & Sörbom, D. (1996a). PRELIS: User's reference guide. Chicago, IL: Scientific Software International.

Jöreskog, K.G., & Sörbom, D. (1996b). LISREL 8: User's reference guide. Chicago, IL: Scientific Software International.

Jöreskog, K.G., & Sörbom, D. (1988a). PRELIS: A preprocessor of LISREL. Mooresville, IN: Scientific Software.

Jöreskog, K.G., & Sörbom, D. (1988b). LISREL 7: A guide to the program and application. Mooresville, IN: Scientific Software.

Lee, S.Y., & Poon, W.Y. (1986). Maximum likelihood estimation of polyserial correlations. Psychometrika, 51, 113-121.

Lee, S.Y., Poon, W.Y. & Bentler, P.M. (1992). Structural equation models with continuous and polytomous variables. Psychometrika, 57, 89-105.

Lee, S.Y., Poon, W.Y. & Bentler, P.M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. British Journal of Mathematical and Statistical Psychology, 48, 339-358.

Muthén, B.O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika, 49, 115-132.

Muthén, B.O. (1987). LISCOMP: Analysis of linear structural equations with a comprehensive measurement model. Theoretical integration and users guide. Mooresville, IN: Scientific Software.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. Psychometrika, 54, 557-585.

Muthén, B.O. (1993). Goodness of fit with categorical and other nonnormal variables. In K.A. Bollen and J.S. Long, (Eds.), Testing structural equation models. Newbury Park, London: Sage Publications.

Muthen, B.O., du Toit, S. H. C. & Spisic, D. (in press). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical outcomes. Psychometrika.

Muthén, B.O. & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. British journal of Mathematical and statistical Psychology, 38, 171-189.

Muthén, L. K., & Muthén, B. O. (1998). Mplus: The comprehensive modeling program for applied researchers: User's guide. Los Angeles, CA: Muthén & Muthén.

Olsson, U. (1979). On the robustness of factor analysis against crude classification of the observations, Multivariate Behavioral Research, 14, 485-500.

Table 1

Thresholds for converting continuous variables to polytomous variables, in z-score form

(and expected percentages in each category for normally distributed underlying variables)

	Categorization	
Number of Categories	1. equal proportions in categories (Uniform distribution)	2. categories approximate Normal (skew=0, kurtosis=0)
2 threshold	0	-
(percentage)	(50 in each)	
3 threshold	-.43 .43	-.97 .97[a]
(percentage)	(33.333 in each)	(16.6 / 66.8 / 16.6)
5 threshold	-.84 -.255 .255 .84	-1.3 -.69 .69 1.3[a]
(percentage)	(20 in each)	(9.7 / 14.8 / 51 / 14.8 / 9.7)
7 threshold	-1.07 -.565 -.18 .18 .565 1.07	-2.05 –1.28 -.47 .47 1.28 2.05[a]
(percentage)	(14.286 in each)	(2 / 8 / 22 / 36 / 22 / 8 / 2)

	Categorization	
Number of Categories	3. large proportion in outer categories: U-shaped (skew=0; kurtosis= -1.7)	4. positively skewed (skew=1.3)
2 threshold	-	.78
(percentage)		(78.2 / 21.8)
3 threshold	-.25 .25	.475 1.5
(percentage)	(40.1 / 19.7/ 40.1)	(68.1 / 25.2 / 6.7)
5 threshold	-.49 -.1 .1 .49	-.05 .77 1.34 1.88 [b]
(percentage)	(31.2 / 14.8 / 8 / 14.8 / 31.2)	(48 / 29.9 / 13.1 / 6 / 3)
7 threshold	-.585 -.20 -.025 .025 .20 .585	-.1 .45 .95 1.45 1.75 2.2
(percentage)	(28 / 14 / 7 / 2 / 7 / 14 / 28)	(46 / 21.3 / 15.5 / 9.8 / 3.3 / 2.6 / 1.4)

threshold: threshold values in z-score units
percentage: expected percentages in each category for normally distributed variables.
[a] values from Bandalos & Enders, 1996
[b] values from Muthén & Kaplan, 1985

Table 2

Mean number of EQS converging replications by sample size (n) and number of categories per variable (CATPRVAR)

Number of EQS Converging Replications out of 200

n	CATPRVAR	Mean	Std. Dev.	Number of Cells
100	2	58.13	20.53	30
	3	46.62	15.34	60
	5	44.08	14.85	60
	7	47.38	16.56	60
	Total	47.76	16.86	210
200	2	116.03	30.57	30
	3	85.57	18.93	60
	5	71.12	18.38	60
	7	72.08	18.73	60
	Total	81.94	25.64	210
500	2	169.10	19.25	30
	3	115.20	19.94	60
	5	90.07	22.66	60
	7	92.80	21.05	60
	Total	109.32	33.78	210
1000	2	188.00	10.44	30
	3	140.55	31.71	60
	5	98.45	26.60	60
	7	100.15	23.93	60
	Total	123.76	40.96	210
5000	2	199.83	00.59	30
	3	151.43	41.23	60
	5	117.67	47.19	60
	7	107.90	25.08	60
	Total	136.26	47.57	210
across all sample sizes	2	146.22	56.02	150
	3	107.87	46.78	300
	5	84.28	37.69	300
	7	84.06	30.46	300
	Total	99.81	46.96	1050

Table 3

Mean and standard deviation of F3,F1 path estimated by LISREL under 5 conditions averaged across all sample sizes

Condition	Mean	Number of cells	Std. Deviation
other	.5043570	770	.0593027
all (9 variables) cat	.5010181	70	.0121333
F1 all cat	.3461238	70	.1038076
F1 and F2 all cat	.4189146	70	.1194977
F1 and F3 all cat	1.0101827	70	.2576756
Total	.5217784	1050	.1659038

Table 4

Mean and ranges of the mean F3,F1 (population value=.5) and F3,F2 (population value=.7) path coefficients for EQS and Mplus by sample size

		F3,F1 = .5		F3,F2 = .7	
N		EQS	Mplus	EQS	Mplus
100	Mean	.4761127	.5107666	.6690429	.7072135
	Range	.14853	.02192	.14247	.02811
200	Mean	.4837527	.5056980	.6796373	.7057644
	Range	.08835	.02304	.08104	.02904
500	Mean	.4841255	.4975284	.6861087	.7019447
	Range	.03844	.00970	.04916	.01446
1000	Mean	.4913530	.4971845	.6964296	.7012248
	Range	.03268	.00636	.02573	.01292
5000	Mean	.5008336	.5002206	.6999663	.6991682
	Range	.01844	.00253	.01117	.00331
Total	Mean	.4872355	.5022796	.6862370	.7030631
	Range	.14853	.03064	.14853	.03064

Table 5

Mean and ranges of the mean F3,F1 (population value=.5) and F3,F2 (population value=.7) path coefficients for EQS and Mplus by number of categorized variables

Number Vars Categorized		F3,F1 = .5		F3,F2 = .7	
		EQS	Mplus	EQS	Mplus
1.00	Mean	.4784956	.5016586	.6767635	.7026572
	Range	.07283	.02558	.06663	.02264
2.00	Mean	.4792122	.5018698	.6782819	.7026859
	Range	.08609	.02898	.07986	.02468
3.00	Mean	.4911555	.5026877	.6885397	.7035170
	Range	.06162	.02982	.06423	.02858
4.00	Mean	.4962009	.5019819	.6961449	.7026643
	Range	.06433	.02490	.07357	.02454
6.00	Mean	.4967817	.5026244	.6966724	.7033035
	Range	.09447	.03019	.11206	.03007
9.00	Mean	.4784786	.5034974	.6789733	.7040984
	Range	.11272	.02883	.11317	.02908
Total	Mean	.4872355	.5022796	.6862370	.7030631
	Range	.14853	.03064	.14247	.03049

Table 6

Mean and ranges of the mean F3,F1 (population value=.5) and F3,F2 (population value=.7) path coefficients for EQS and Mplus by Endogenous – Exogenous factors with categorized variables

Endogenous and Exogenous Factors		F3,F1 = .5		F3,F2 = .7	
		EQS	Mplus	EQS	Mplus
F1 cat - 1 exo	Mean	.4851010	.4999302	.6803832	.7002155
	Range	.07672	.02176	.06856	.00471
F3 cat - 1 endo	Mean	.4810175	.5041717	.6819147	.7050932
	Range	.08975	.02713	.07209	.02782
F1 and F2 cat –	Mean	.4895560	.4998972	.6932207	.7004130
2 exo	Range	.08742	.02236	.09205	.01232
F1 and F3 cat –	Mean	.4907531	.5037099	.6885128	.7049770
1 exo 1 endo	Range	.08721	.03019	.09203	.02744
all factors -	Mean	.4897498	.5036890	.6871533	.7046169
have cat vars	Range	.14853	.02883	.11891	.02908
Total	Mean	.4872355	.5022796	.6862370	.7030631
	Range	.14853	.03064	.14247	.03049

Table 7

Mean and ranges of the mean F3,F1 (population value=.5) and F3,F2 (population value=.7) path coefficients for EQS and Mplus by number of categories per variable

Number of Categories per Variable		F3,F1 = .5		F3,F2 = .7	
		EQS	Mplus	EQS	Mplus
2	Mean	.4864167	.5023774	.6838662	.7028560
	Range	.14853	.03019	.12928	.02172
3	Mean	.4879672	.5027688	.6877767	.7041707
	Range	.08122	.03028	.09228	.02891
5	Mean	.4871067	.5020636	.6867040	.7025443
	Range	.09982	.02369	.09173	.01495
7	Mean	.4870419	.5019575	.6854155	.7025779
	Range	.09690	.02384	.10522	.01467
Total	Mean	.4872355	.5022796	.6862370	.7030631
	Range	.14853	.03064	.14247	.03049

Table 8

Mean and ranges of the mean F3,F1 (population value=.5) and F3,F2 (population value=.7) path coefficients for EQS and Mplus by categorized distribution due to threshold location

Categorized Distribution		F3,F1 = .5		F3,F2 = .7	
		EQS	Mplus	EQS	Mplus
Uniform	Mean	.4866114	.5025610	.6856734	.7021789
	Range	.09006	.02835	.08855	.01791
Normal	Mean	.4883081	.5018160	.6872872	.7033227
	Range	.08932	.02270	.10063	.01745
Platykurtic	Mean	.4875751	.5026613	.6867199	.7021988
	Range	.09528	.02353	.08373	.01284
+Skewed	Mean	.4868003	.5020596	.6856506	.7044009
	Range	.14853	.03064	.13336	.02864
Total	Mean	.4872355	.5022796	.6862370	.7030631
	Range	.14853	.03064	.14247	.03049

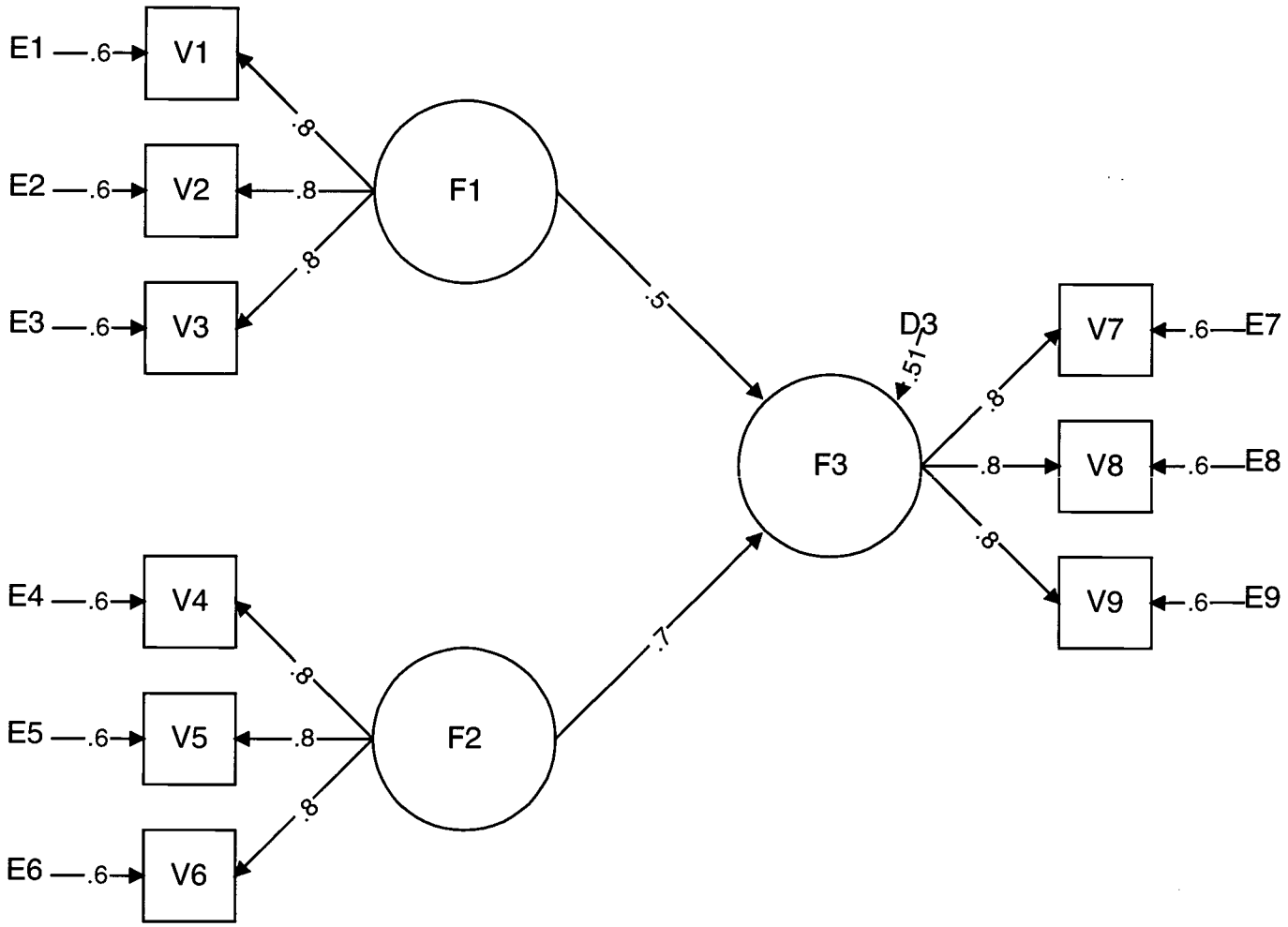
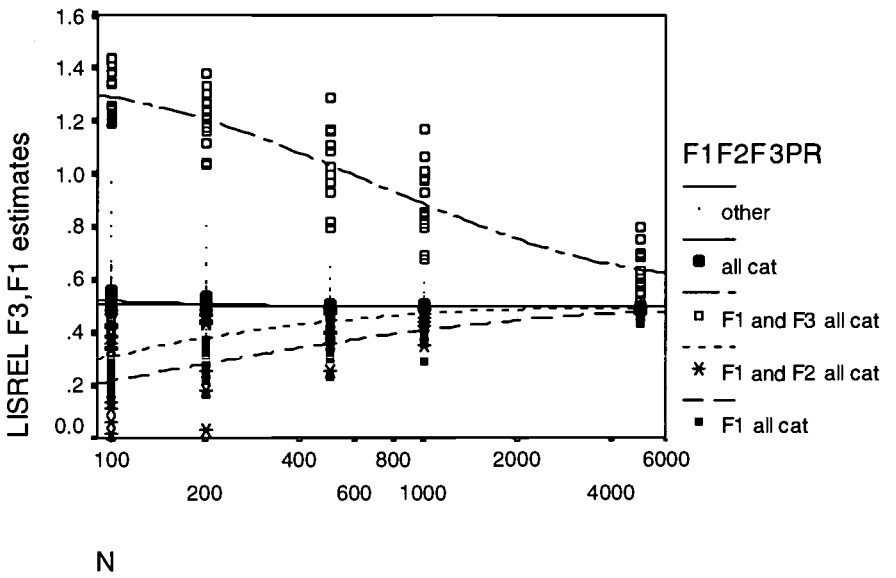


Figure 1. Population Model

(a)



(b)

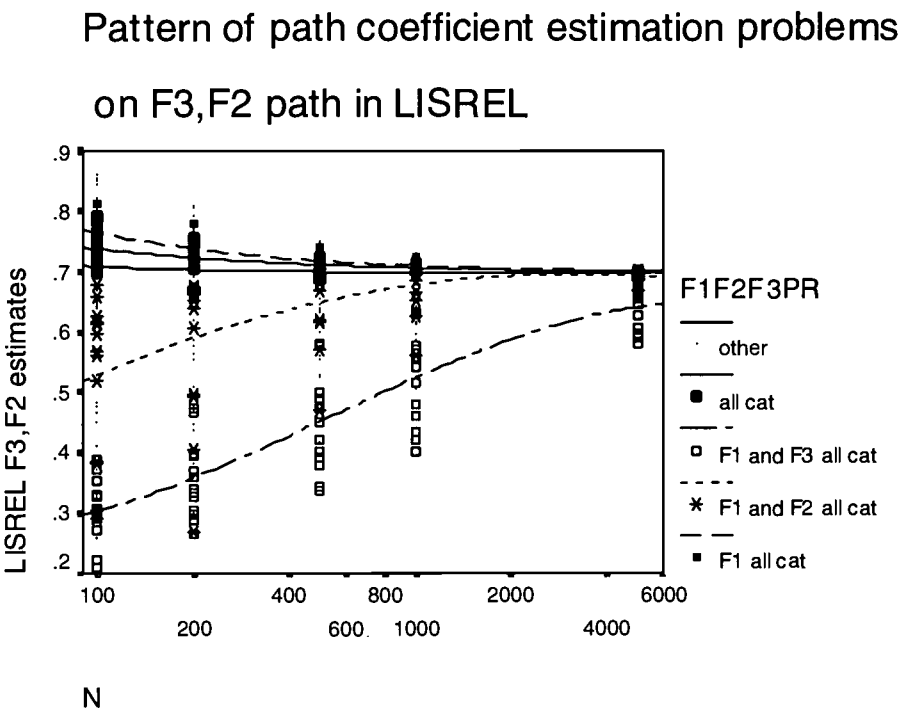
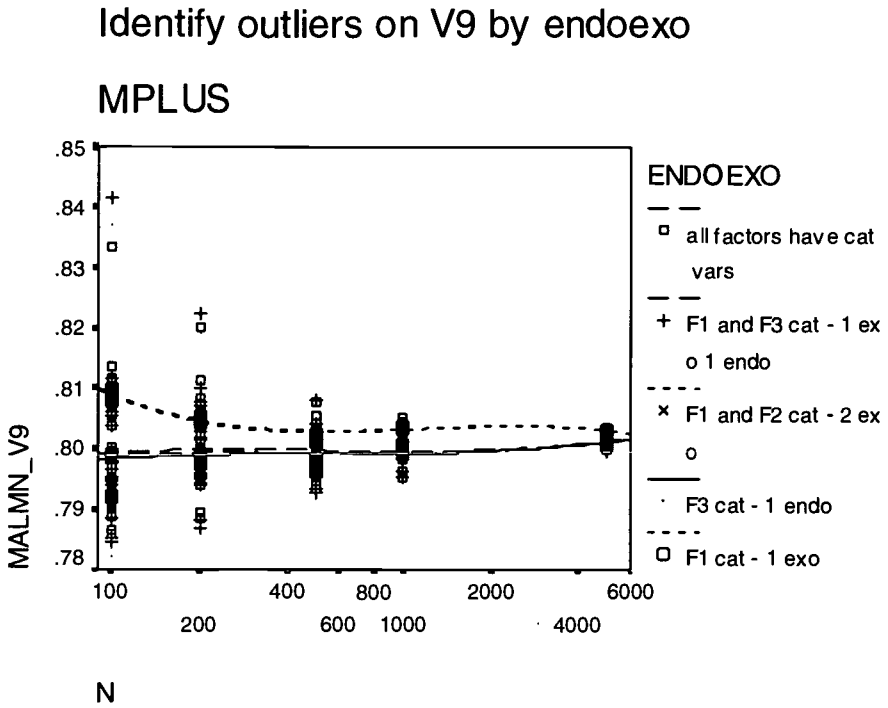


Figure 2. Pattern of path coefficient estimation problems in LISREL
(a) on the F3,F1 path
(b) on the F3,F2 path

(a)



(b)

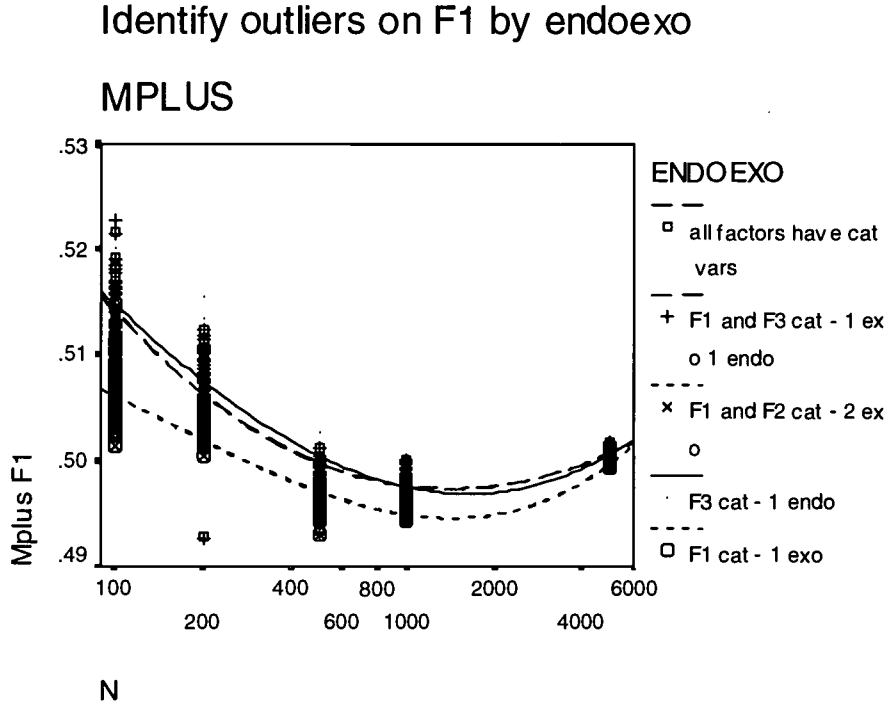
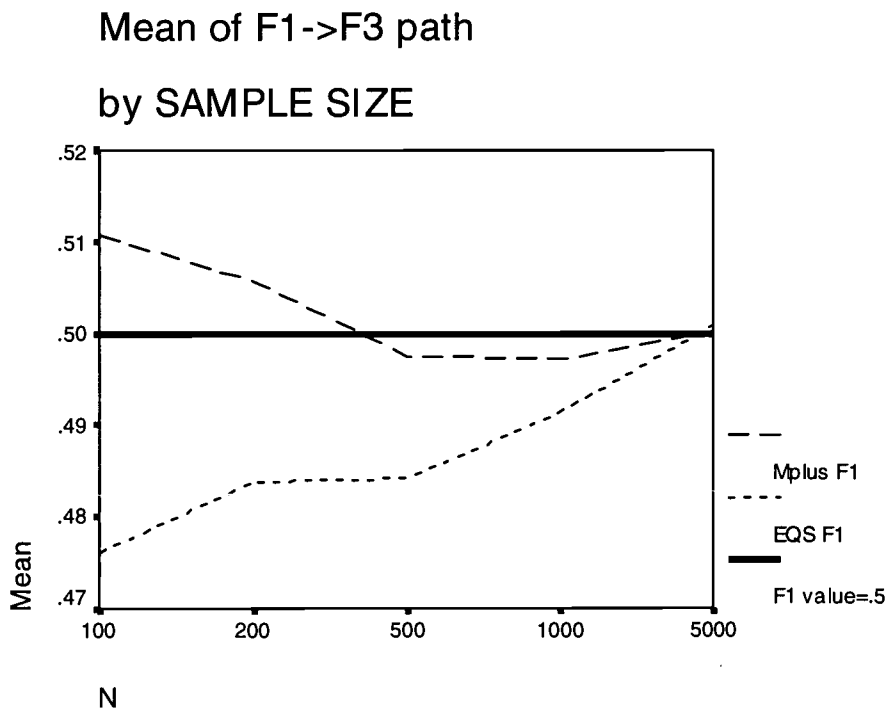


Figure 3.
(a) Estimation of factor loadings in Mplus are slightly closer to the true value for Variables in conditions where the factor they load contains at least one categorized Variable.
(b) Estimation of the path coefficients in Mplus is slightly higher in conditions where V7, which is the scale factor for F3, is a categorized Variable.

(a)



(b)

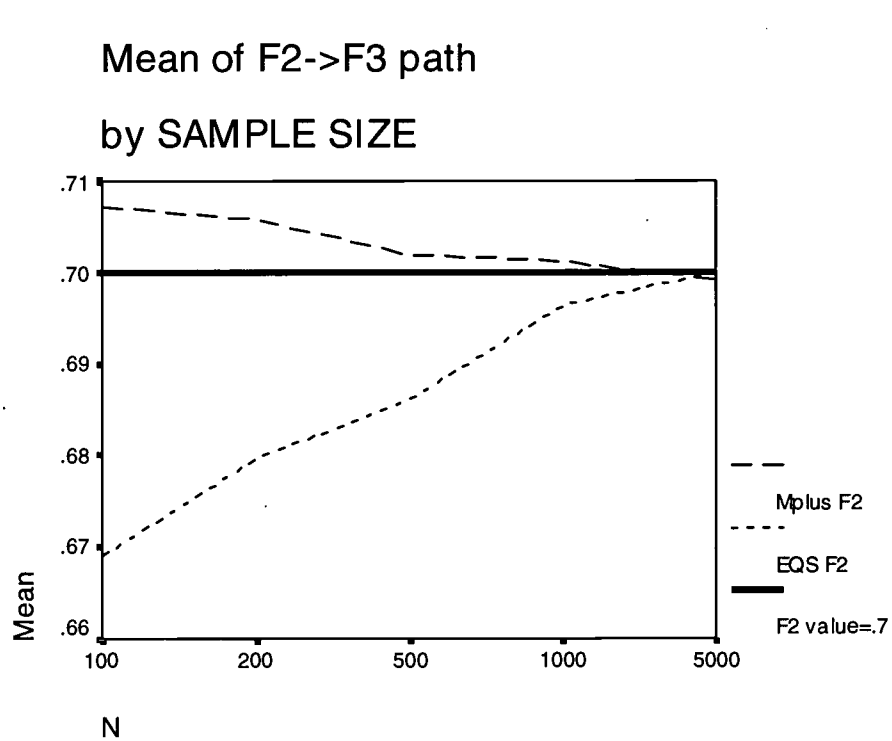
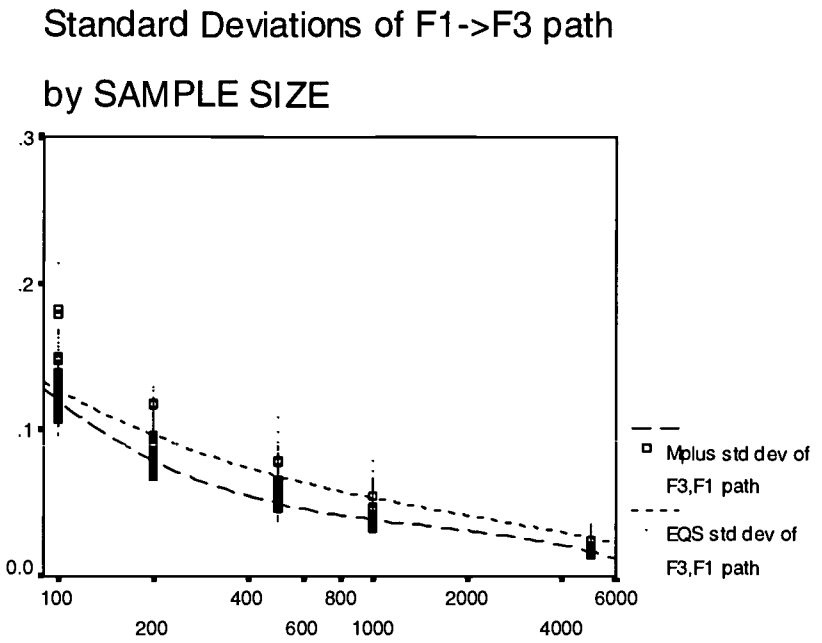


Figure 4.

(a) Estimation of the path coefficient F3,F1 in Mplus are slightly closer to the population value on average than the estimates from EQS.

(b) Estimation of the path coefficient F3,F2 in Mplus are slightly closer to the population value on average than the estimates from EQS.

(a)



(b)

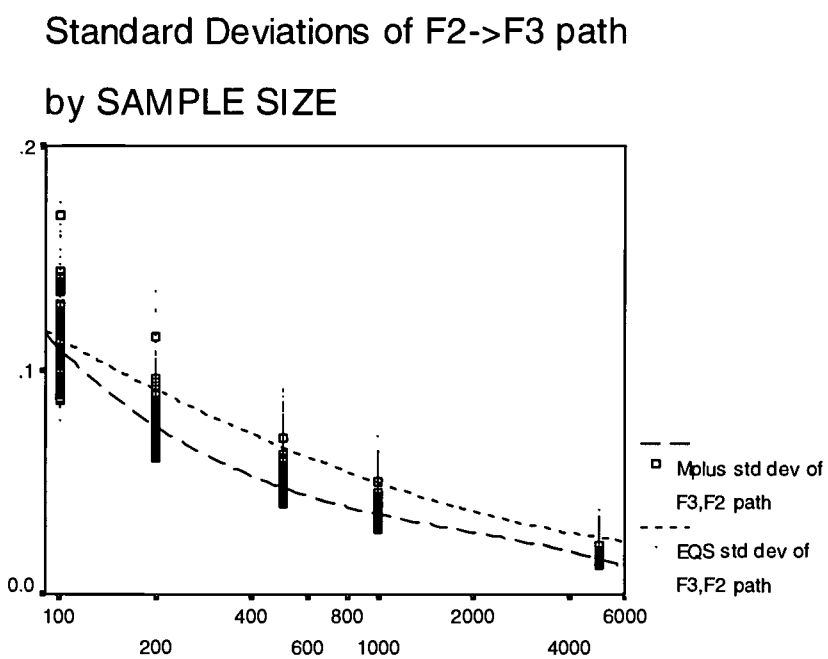


Figure 5.
(a) Estimation of path coefficient F3,F1 in Mplus are slightly closer to the population value than the estimates from EQS.
(b) Estimation of path coefficient F3,F2 in Mplus are slightly closer to the population value than the estimates from EQS.

AERA

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

TM031636

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: A Comparison of methods for structural modeling with polytomous and continuous variables.

Author(s): Ilona Berkovits + Gregory R. Hancock

Corporate Source: University of Maryland

Publication Date: April 2000

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☑

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Ilona Berkovits

Printed Name/Position/Title: Ilona Berkovits / Graduate Student

Organization/Address: University of Maryland, College Park, MD 20742

Telephone: 301-405-3624

FAX: 301-314-9245

E-Mail Address: ilona@wam.umd.edu

Date: 6/26/00

(over)

Clearinghouse on Assessment and Evaluation

University of Maryland
1129 Shriver Laboratory
College Park, MD 20742-5701

Tel: (800) 464-3742
(301) 405-7449
FAX: (301) 405-8134
ericae@ericae.net
http://ericae.net

March 2000

Dear AERA Presenter,

Congratulations on being a presenter at AERA. The ERIC Clearinghouse on Assessment and Evaluation would like you to contribute to ERIC by providing us with a written copy of your presentation. Submitting your paper to ERIC ensures a wider audience by making it available to members of the education community who could not attend your session or this year's conference.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed, electronic, and internet versions of *RIE*. The paper will be available **full-text, on demand through the ERIC Document Reproduction Service** and through the microfiche collections housed at libraries around the world.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse and you will be notified if your paper meets ERIC's criteria. Documents are reviewed for contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our processing of your paper at **http://ericae.net.**

To disseminate your work through ERIC, you need to sign the reproduction release form on the back of this letter and include it with **two** copies of your paper. You can drop of the copies of your paper and reproduction release form at the ERIC booth (223) or mail to our attention at the address below. **If you have not submitted your 1999 Conference paper please send today or drop it off at the booth with a Reproduction Release Form.** Please feel free to copy the form for future or additional submissions.

Mail to: AERA 2000/ERIC Acquisitions
The University of Maryland
1129 Shriver Lab
College Park, MD 20742

Sincerely,

Lawrence M. Rudner

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

ERIC/AE is a project of the Department of Measurement, Statistics and Evaluation at the College of Education, University of Maryland.